

基因表達規劃法之動態交易策略設計 研究 - 以臺灣股市為例

林青峰・林湘霖・李明祐*

(收稿日期：113 年 02 月 05 日；第一次修正：113 年 06 月 16 日；
接受刊登：113 年 07 月 29 日)

摘要

隨著金融科技發展，人工智慧被應用於各類金融商品上，如何從複雜多變的金融市場找出一套能夠獲利的交易策略是一項挑戰。本研究採用基因表達規劃法 (GEP) 從過去歷史資料中搜尋有效的交易策略，以交易訊號、資金管理與風險控管三種基因構建一套交易模型。此外，隨機森林 (RF) 與 LGBM 兩種決策樹方法，在每一期交易標的中動態更新投資組合，期望透過此種方法提高交易模型的獲利與降低交易時的風險。實驗結果發現，藉由動態選股結合 GEP 交易模型（基本面與技術面）的方法，相較於純 GEP 模型（技術面）與選股模型（基本面），確實能有效的提升 GEP 交易模型各期的交易績效，如獲利、勝率與風險等。

關鍵詞彙：隨機森林、LGBM、基因表達規劃法、交易策略

壹·導論

一、研究背景與動機

隨著科技的發展，人工智慧被應用在各個領域中，其中在金融方面，關於股票的趨勢預測一直以來都是非常熱門的研究議題。這歸因於電腦基礎運算能力的提升與網路資訊的普及，人們能更快速透過網路取得股票的相關資訊，進而分析找出未來具有潛力的標的。在分析股票中較常被使用的有基本面分析與技術面分析，然而股價是否真的能夠被預測，在不同的文獻中持有不同的論點。根據效率市場假說 (Efficient Market Hypothesis, EMH)，未來股價的走勢是呈現隨機漫步且無法被預測 (Fama, 1970)。在弱勢效率市場中，股票價格已充分反映歷史資訊，投資人無法再透過技術分析來獲取超額報酬；在半強式效率市場中，股價反映了已公開的股票資訊；而在強式效率市場中，股票價格已充分

* 作者簡介：林青峰，輔仁大學資訊管理學系助理教授；林湘霖，輔仁大學資訊管理學系助理教授（通訊作者）；李明祐，輔仁大學資訊管理學系碩士生。

反映了股票所有的資訊包括未公開的情報，任何投資者都無法在強式效率市場中獲取超額報酬。這歸因於其假設股市中的投資者皆時刻是理性的，且股票公開的相關資訊都能夠即時的反應在股價上。

然而有些學者則認為股市中的投資者並非全然是理性的，像是 Lo (2004) 提出適應市場假說 (Adaptive Market Hypothesis, AMH)，補充了效率市場假說不足的部分。其主張股市中的投資者行為並非時刻處於理性狀態，而是介於理性與非理性之間，投資者會吸取過去的教訓，並根據相關投資經驗對未來做出預測，而 AMH 也在多篇研究中被驗證比 EMH 更能解釋股票市場上效率的變化 (Lim & Brooks, 2011; Urquhart & Hudson, 2013; Lekhal & El Oubani, 2020)。在實際進行交易時，就如同 AMH 所解釋的，當投資者進行交易時會依自身交易經驗去做判斷，而這個過程往往會產生過度自信 (Over Confidence) 與處置效應 (Disposition Effect) 等問題，進而提高交易的風險並影響整體投資績效 (Barber & Odean, 2000)。如透過如機械式交易或程式交易則能避免人性問題，像是利用程式控制停損停利和資金管理，或是使用回測平台評估策略的優劣，從而使交易策略更加穩定與達到控制風險的目的。

交易策略評估與最佳化之研發方法可分為兩種，其一為將投資者過去的經驗量化後透過科學方法驗證，每次回測後進行修正，最終找出一個適合市場的交易策略；其二是透過人工智慧 (Artificial Intelligence, AI) 從而分析過去大量的交易資料，並從中找出樣式、規則或訊號後再進行交易 (李佳儒, 2010)。一些研究已證實機器學習 (Machine Learning) 和深度學習 (Deep Learning) 等預測模型，在股票預測之準確度和處理大量資料的效率方面，是明顯優於傳統統計預測模型的 (Bustos & Pomares-Quimbaya, 2020; Rundo et al., 2019)，尤其在產生股票交易之規則上，遺傳演算法 (Genetic Algorithm, GA)、遺傳規劃法 (Genetic Programming, GP) 與基因表達規劃法 (Gene Expression Programming, GEP)，皆提供了不錯的解決方案，其能從過去數據中探勘最佳化策略且產生的交易規則也具有可解釋性，讓投資決策者清楚演算法運作邏輯，並能在獲取穩定的超額報酬下同時降低風險 (Allen, 1999; 陳帝豪, 2012; 韋承佑, 2014; Michell & Kristjanpoller, 2020)。

GEP 演算法結合了 GA 與 GP 兩種演算法的特點，具有前者之線性編碼在遺傳操作時的效率，以及後者以表達樹定義問題來傳遞高度可解釋性結果予分析者之優點，能在處理複雜問題上同時保有演化時的效率與成效 (Ferreira, 2006)。然而，儘管 GEP 有許多優點，但其在某些地方仍有其不足之處：第一，欲提升 GEP 的性能，須先進行複雜的參數設計；第二，需使用不同的機器學

習模型來輔助 GEP，例如構建代理模型與特徵工程降維等；其三，不易設計與優化用於處理多任務或是存在多個目標問題之 GEP 框架 (Zhong et al., 2016; 2017)。

二、研究目的

主動投資包含選股與擇時兩種方法，而目前關於以機器學習作為股價預測的模型中，以單使用技術面的模型佔多數，其次是只依基本面 (Nti et al., 2019)。近年來，使用基本面與技術面混合模型之研究正逐漸增加，原因在於同時考慮內外部因子更能提供精確的預測能力。此外，陳帝豪 (2012) 與韋承佑 (2014) 的動態交易 GEP 研究中發現，當訓練期與測試期的趨勢不一致時，GEP 交易模型往往較容易出現虧損情況。為避免該情況發生，本研究欲透過動態選股的方式改良 GEP 交易模型的限制，透過染色體中基因的設計以及選定適應函數，藉此協助投資者有效率地從龐大的資料中挖掘出未來具有上漲可能的股票，並進行股票之投資組合規劃，提供有效的交易策略。

此外，本研究將以機器學習中之隨機森林 (Random Forest, RF) 與 LightGBM (LGBM) 兩種決策樹 (Decision Tree) 演算法，分別訓練一個選股與擇時的模型，相較於其他機器學習算法與深度學習模型無法得知背後交易邏輯，決策樹具有更高的可解釋性，且能清楚了解背後運作規則。另一方面，透過集成方法 (Ensemble Method) 的決策樹，能具有更優越的分類能力。

根據上述欲採用之演算方法，本研究將使用基本面與技術面兩種特徵資料，分別構建一套選股及擇時的交易模型，以選股的方式從台灣上市上櫃公司中找出未來上漲的股票，並透過擇時的方式從這些個股中找出最適的交易規則，主要研究目的條列如下：

1. 使用各種選股因子於集成決策樹模型，找出未來有較大上漲可能的股票，探討選股模型所產生的標的能否提升 GEP 模型的交易績效，以及解決 GEP 交易模型在趨勢不同時的弱點。
2. 將選股池所挑選出的股票標的，藉由移動窗格 (Walk Forward) 的方式分割出包含各種走勢的時間序列資料，再利用 GEP 以技術指標動態規劃出能獲取穩定超額報酬的交易策略，並從每一期的規劃中各別評估其績效與策略的穩定性。

貳·文獻探討

一、證券交易

股票是公司以籌措資金為目的所發行的一種有價證券，持有一間公司的股票能代表你擁有公司部分的資產。當持有股票時從中賺取報酬的方式主要有兩類方式，第一種是透過公司所配發的股利 (Dividend) 賺取收益，其中股利又可分為以現金發放的現金股利與以股票發放的股票股利兩種。第二種是從股價隨時間的波動中賺取價差從中獲得資本利得 (Capital Gains)。而在台灣進行股票投資交易時，能藉由幾種分析技術判斷是否購買該檔標的，主要能分為基本面與技術面，以下為兩者的介紹。

(一) 基本面分析

基本面分析係當投資人進行交易時，可以參考總體經濟景氣、個別產業狀況以及每間公司的營運狀況，藉此預測證券的真實價值與未來的投資收益 (鍾祥龍，2008)。其中公司的營運能透過各家公司所提供的財務報表進行評估，找出價值被低估的公司並進行投資。財務報表主要能分為現金流量表 (Cash Flow Statement)、資產負債表 (Balance Sheet) 與損益表 (Income Statement)，在選股方面常利用以上三種報表挑選公司體質優秀的標的，而為了容易挑選衍生出各種財務指標，Fama & French (2015) 提出五因子定價模型包含市場報酬率 (Market Return)、市值 (Small Minus Big, SMB)、帳面市值比 (High Minus Low, HML)、獲利能力 (Robust Minus Weak, RMW) 與投資水平 (Conservative Minus Aggressive, CMA) 等五種風險因子，五因子模型相較於三因子模型更具解釋力。

(二) 技術面分析

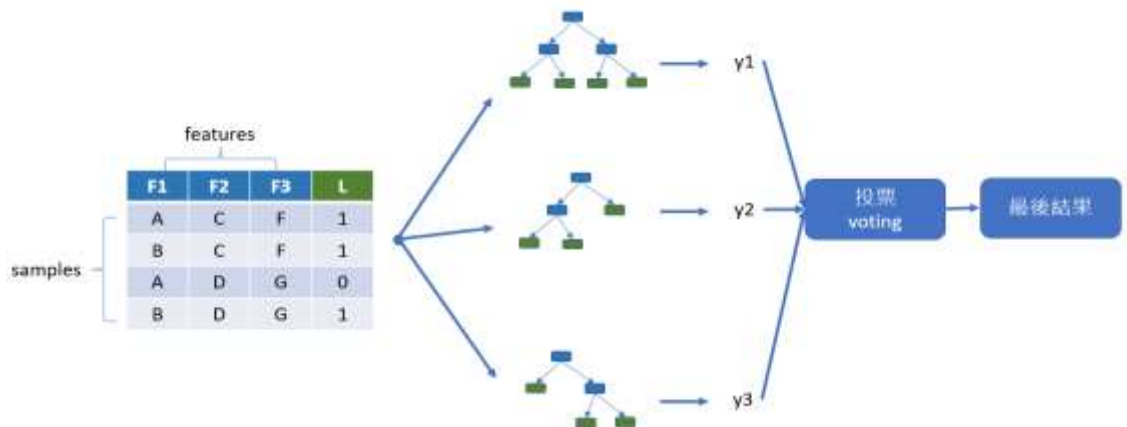
技術面分析又可稱為市場分析或是內部分析，藉由研究過去股價和交易量的變動，試圖找出未來股價變動的方向。技術分析師相信，市場只有 10% 的理性，加上 90% 的心理因素，雖然圖形只能顯示出市場參與者過去的行為，但藉由研究其他人過去的行為，就能夠預測群眾未來可能的走向 (張朝棟，2011)。在一些研究中也顯示使用技術分析能夠為投資者帶來超額報酬，Ko 等人 (2014) 也認為除了價值投資外使用一些技術指標能幫助投資者把握進出場的時機。張清良 (2008) 使用移動平均線 (Moving Average, MA)、指數平滑異同移動平均線 (Moving Average Convergence & Divergence, MACD)、相對強弱指

標 (Relative Strength Index, RSI)、乖離率 (Bias Ratio, BIAS)、動量指標 (Momentum Index, MTM)等八種指標於摩根台指成分股中評測指標的表現，研究結果顯示 MACD、MA 與 MTM 指標績效較佳，BIAS 指標的表現最差，並指出 MACD、MA 與 MTM 具有預測能力但是無法適用於全部類股，且在同類股中所適用的交易規則能有些許不同。

二、決策樹

(一) 隨機森林

隨機森林 (Random Forest, RF) 決策樹由 Breiman (2001)所提出，其原理是將傳統決策樹與集成學習中的袋裝法 (Bagging) 結合，能使模型提高準確度並避免過度擬合 (Overfitting) 之情況，其概念如下圖一所示。首先將原始資料集透過 Bagging 的方式隨機分割樣本與特徵資料，並由分割出的資料訓練出個別決策樹，藉由這種方法建構出多顆含有不同資料集的決策樹形成森林，再將實際的資料輸入至每顆決策樹中，最後透過投票 (Voting) 的方式輸出最終的結果。



圖一 隨機森林流程圖

(二) LGBM

LGBM (Light Gradient Boosting Machine) 係 Ke 等人 (2017) 所提出的一種新型梯度提升樹，具有高速、高準確度、高穩定性等優點。本質上是一種增強型的梯度提升樹，可用於回歸與分類問題；與之前的 GBT 相比不同的地方在於，LGBM 採用兩種不同的方式提高模型的準確性與效率，分別為基於梯度的單邊抽樣 (Gradient-based One-Side Sampling, GOSS) 與獨立特徵合併

(Exclusive Feature Bundling, EFB) 兩種技術;使用 GOSS 在於將重要的特徵保留並剔除梯度較小的資料,而 EFB 在於將多個特徵合併達到特徵降維之目的。

(三) 決策樹相關文獻

有關使用 RF 演算法應用於股票預測之相關研究文獻, Khaidem (2016) 以 RF 結合 RSI、KD、威廉等多項技術指標預測股票市場, 研究結果顯示其所訓練的模型在預測股票未來走勢上可達 85 以上的準確度; 而 Tan 等人 (2019) 同樣使用 RF 作為選股模型並應用於中國股票市場, 並藉由移動視窗評估其選股模型的穩健性。研究結果顯示技術指標的動量策略模型產生的盈利能力高過基本面模型; Hsu (2021) 使用 RF 作為選股模型, 並選擇基本面中六個主要因子 (分別為規模、品質、價值、動能、高股息與低波動因子), 結合技術分析與籌碼分析作為模型的特徵。研究結果顯示在不考慮交易成本的情況下, 其模型之績效可勝過多數指數之買入持有策略。

而亦有不少過往研究使用了 LGBM 技術相關文獻, 對各種金融市場進行預測模型之建立。Chen 等人 (2020) 即使用 LGBM、支援向量機 (SVM)、RF 與深度神經網路 (DNN) 等多種機器學習技術, 建立中國股市選股模型。研究結果顯示, LGBM 無論是準確度或是穩定性, 都比其他模型表現得更出色; Padhi 等人 (2021) 同樣使用 LGBM 來評估四個不同國家的指數績效, 研究結果同樣顯示, 該決策樹演算法在模型的訓練期與測試期的績效差異度不大, 相較於過往追求報酬率的模型更適合提供給交易者使用; 而 Sun 等人 (2021) 則將 LGBM 應用在加密貨幣市場, 透過分析每日價格變化預測未來加密貨幣的漲跌。結果顯示該方法非常適合對這類金融商品市場進行中期 (2 週) 的預測。

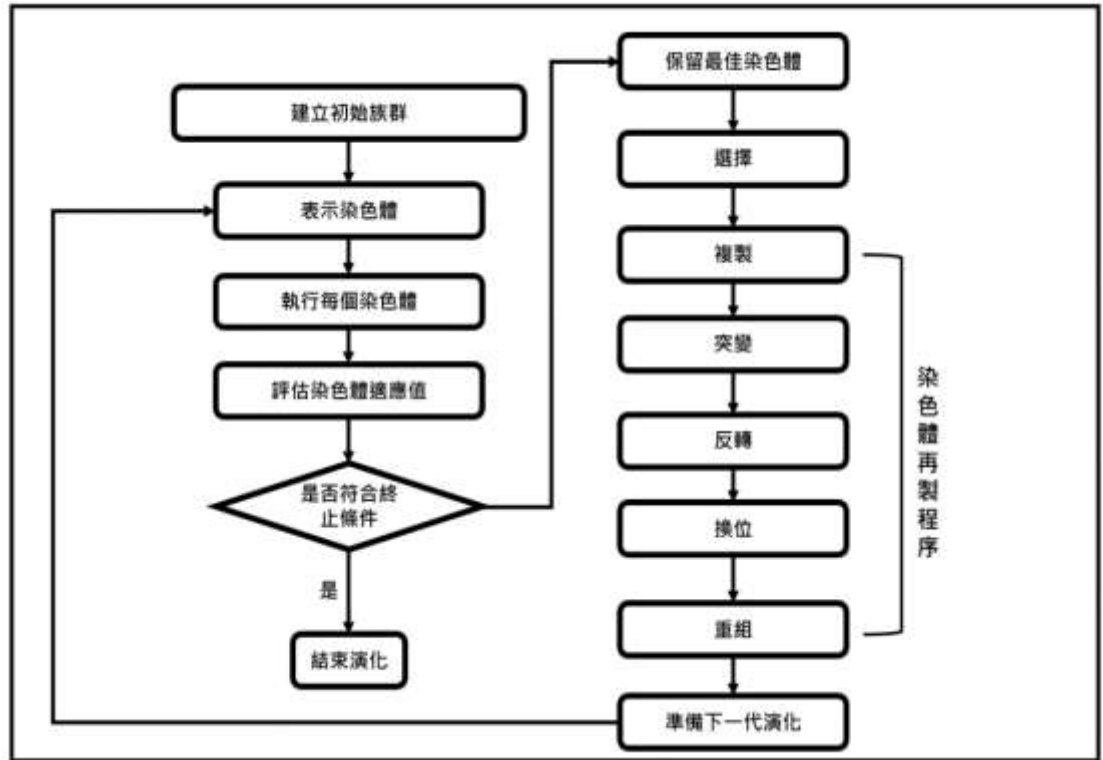
三、基因表達規劃法

基因表達規劃法 (Gene Expression Programming, GEP) 由 Ferreira (2001) 所提出, 結合遺傳演算法 (Genetic Algorithm, GA) 與遺傳規劃法 (Genetic Programming, GP) 兩種演算法的特點, 具有 GA 固定長度編碼與 GP 的非線性解析樹表示方式, 能夠在處理複雜問題上同時並保有運算效率。

(一) 演算法流程

首先會以隨機的方式建立一個初始的族群, 族群中會有多個染色體, 每個染色體都代表一個解決方案, 接著將族群中各個染色體由基因型 (Genotype) 轉換成以表達樹呈現之表現型 (Phenotype), 執行每個表達樹並評估其適應值,

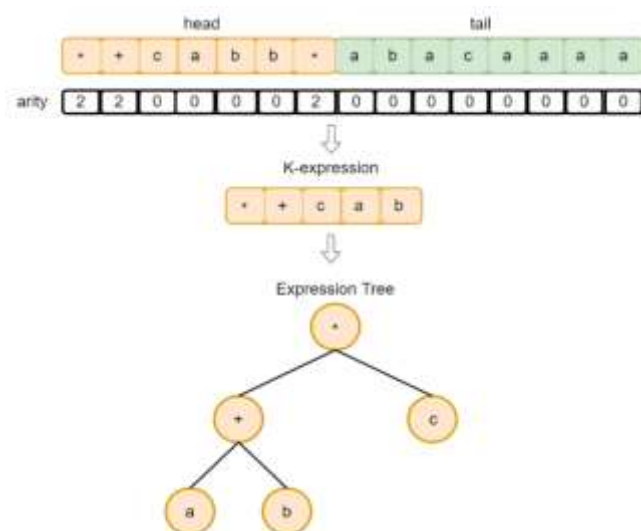
透過選擇的機制能夠將適應值高的染色體保留並淘汰表現不佳的個體，保留下來的個體會經過染色體再製程序，包含複製、突變、反轉、換位與重組將基因保留至下一代中，然後進行一次迭代，最後當達到所設定的最大迭代次數或是適應值標準終止演化並提取最佳染色體。流程圖如圖二所示。



圖二 基因表達規劃法流程圖 (Ferreira, 2001)

(二) 基因編碼方式

在基因表達規劃法中染色體會具有基因型與表現型兩種表現方式，基因型是指 GEP 在演化時所使用的線性基因字串，用於染色體再製工作，表現型則是線性基因所轉換成的表達樹，用於讀取基因規則並以適應函數評估其適應值。在 GEP 的染色體中並非每個基因都會被轉換成表達樹，是先由開放閱讀框架 (Open Reading Frame, ORF) 找尋有效的基因長度，此方法在 GEP 中又可稱為 K 表達式。轉換表達樹的規則是由上至下，由左至右依序地將有效基因填入至表達樹中，每個基因是透過其分支數量判斷是否為終端節點，若當前節點為終端節點則表達樹停止擴展，以圖三所示。



圖三 K 表達式示意圖

(三) 基因表達規劃法相關文獻

有關使用 GEP 應用於股票市場交易策略之相關研究文獻，陳帝豪 (2012) 使用最佳資源配置尋找 GEP 最佳參數，並透過 GEP-ADF 的編碼方式，將多種技術指標分配至個別自動定義函數 (ADF) 中演化，再藉由動態調整技術找出指標臨界值，最後以投票方式來判別買賣訊號，研究結果也發現此作法確實能提高投資獲利；而韋承佑 (2014) 結合 GEP 與決策樹，從歷史資料中探勘交易策略，並以 ADF 的方式將常見的交易規則嵌入至樹節點中，並在不同盤勢中動態調整交易規則的標準。研究結果顯示，其預測模型在上漲趨勢中能確保獲得利潤且在下跌走勢時亦能降低損失風險；詹雅婷 (2015) 將基本面、籌碼面與公司治理三類因子共 33 項指標，透過 GEP 篩選出具有潛力的指標，並結合遺傳演算法 (GA) 進行資金管理配置。研究結果顯示，GEP-GA 資金管理選股模型有效提升報酬率並具有風險管理機制。

在使用 GEP 作為建立股票預測模型的文獻中，有不少研究係特別關注於技術面。陳姿穎 (2016) 所提出之模糊 GEP 模型即是用以解決技術指標鈍化的問題，其基因編碼具有買賣訊號、停損停利與資金加減碼三種功能，並透過隨機數值常數動態調整指標天數週期，藉此在訓練期間有效找出最佳的獲利規劃；呂冠融 (2018) 則以長短期記憶網路 (LSTM) 結合技術面預測台灣中型 100 成分股中，具有獲利潛力的個股，並透過 GEP 所訓練的交易策略來獲取超額報酬，研究結果顯示在每個實驗期中皆獲得不錯的績效；黃賜傳 (2018) 以 GEP 模型結合模糊理論，並採用五種技術指標作為 GEP 基因節點進行動態

標竿設計，實驗結果發現其模型之績效可勝過諸多同期買入持有策略；而李家璋 (2020) 也以 GEP 作為特徵工程篩選指標技術，並透過移動窗格讓所分割的視窗各別演化，找出每個區間前五名被選中的指標。後將萃取後的技術指標當作特徵值來訓練股價預測模型。

四、文獻總結

從過去的文獻能得知，基本面與技術面分析都有其支持者，這些分析理論使用在不同的國家、市場與商品得出的研究結果皆有所不同，有些市場適合使用基本分析而有些適合技術面，甚至是兩者間結合應用。其中，技術面分析在較短天期的效果較好，而長時間的趨勢大部分的投資者較偏重於使用基本面分析 (Hsu, 2021)。

隨著各種分析指標的衍生需要處理的資料越來越龐大，而在處理巨量資料時人工智慧是相較於傳統統計模型來的表現更好，在預測方法中集成式的決策樹模型已被多篇研究證實具有預測能力。此外，從過往研究也發現，GEP 強大的規劃求解能力能夠從股市中探勘出一個有效而穩定的交易策略，例如，可透過設計不同的染色體基因擴展模型的功能，在基本面選股、擇時交易、資金管理或是停損停利等多方面進行預測模型建模。綜合過去在以 GEP 作為擇時交易模型的文獻中能得知以下幾點：

1. 從過去的文獻中得知技術指標常被用於 GEP 交易模型中，其中較常見且重要的技術指標包含隨機指標 (KD)、相對強弱指標 (RSI)、威廉指標 (WMS%R) 與移動平均線 (MA) 等。在設計 GEP 交易模型的屬性節點時，加入適當的技術指標，能使模型具有較佳的求解能力。
2. 夏普指標 (Sharp Ratio)、絕對報酬率 (Return on Investment, ROI) 與淨利 (Net Profit) 作為 GEP 的適應函數，在過去研究中皆已被證實能有效找出交易策略。尤其使用 ROI 作為適應函數，其模型績效更是勝過同期的買入持有策略。
3. 部分研究發現，以移動窗格方式切割的時間序列資料作為訓練並評估 GEP 模型其結果發現，以長期交易來看，模型能夠穩定在各個期間中獲取穩定的超額報酬，且在投資標的處於熊市時，也能降低交易風險。但模型如處在訓練與測試期走勢不一致時，當期的績效表現往往會產生虧損，此為 GEP 模型的一大限制。

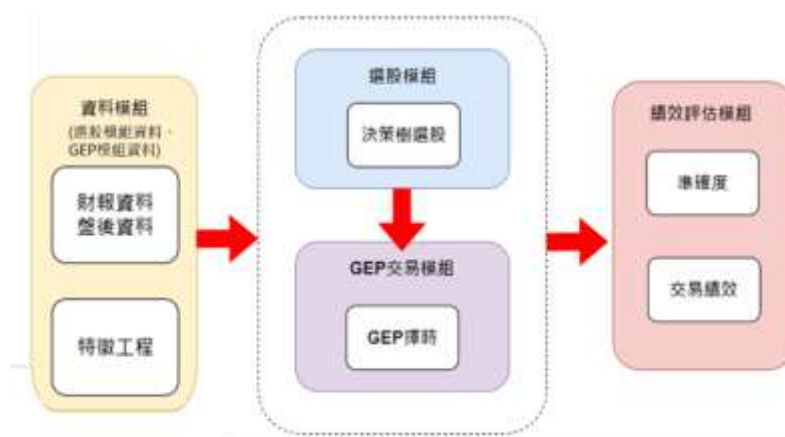
基於上述，本研究以提升 GEP 模型於訓練期與測試期相異期間的表現做為目的，透過將 GEP 交易模型結合決策樹選股方法，藉此避免交易時趨勢不一致所帶來的虧損風險。

參.研究設計

本研究使用兩種決策樹演算法與基因表達規劃法 (GEP) 分別作為選股以及擇時的交易模型。在決策樹方面透過選股因子預測出未來具有上漲可能的股票作為交易標的，再將每一期中決策樹所預測的最佳標的透過 GEP 交易模型找出其最佳的交易策略，期望能透過增加選股模型與多元交易系統達到提升報酬與減少風險的目的。

一、研究架構

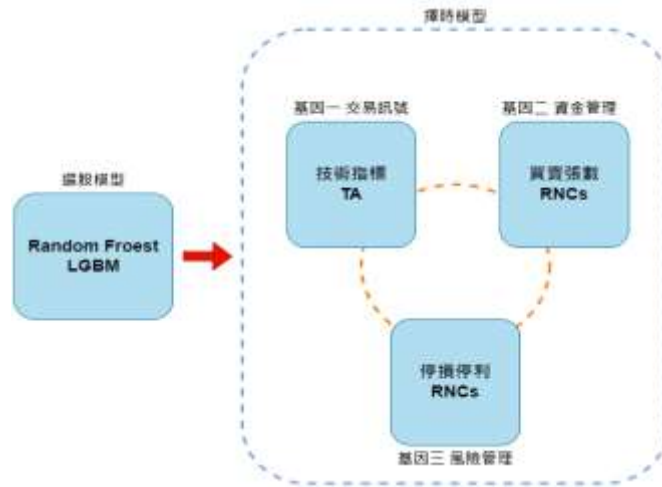
本研究架構主要分成資料、選股、GEP 交易與績效評估四個模組。資料模組內包含挑選之企業的財務報表以及每日盤後資料，作為後續決策樹與 GEP 模型進行建模之使用。接著從選股模組中挑選適合的股票標的再以 GEP 模組進行交易，最後透過績效評估模組評估模型於訓練期與測試期的模型表現與交易策略績效。研究架構如圖四所示。



圖四 研究架構圖

在訓練與測試（交易）時會以選股模組與 GEP 模組相互合作，由選股模組挑選出未來具有上漲可能的股票，提供給 GEP 模組進行交易。透過 GEP 交

易模組來判斷各檔股票合適的進出場點、買賣的張數與停損停利時機，藉此完成一次完整的交易。其架構圖如圖五所示。



圖五 選股模組與 GEP 模組

二、資料模組

本研究之資料模組蒐集了幾間台灣上市上櫃公司的財務報表以及其每日之盤後資料，並分別預先對這些原始資料進行前處理和轉換等特徵工程後，萃取所需之各項指標並建立選股模組資料與 GEP 模組資料，以利後續提供給決策樹與 GEP 模型進行建模。

(一) 選股模組資料

選股模型資料中的特徵選自各公司的財務報表資料，並參考 Hsu (2021) 對投資人在進行交易標的基本面分析時，預測證券真實價值與未來投資收益常使用之六因子選股模型，以規模、品質、價值、動能、高股息與低波動等因子作為選股指標特徵，彙整如表一。

表一 選股指標表

編號	因子	指標名稱
1	規模因子	市值(Market Value)
2	品質因子	每股盈餘(EPS)、淨利率、ROA、ROE
3	動能因子	資產增長率、營收增長率、息稅折舊攤銷前盈餘增長率
4	價值因子	本益比(P/E)、股價淨值比(P/B)
5	波動因子	每日收益標準差
6	股利因子	股息報酬率

(二) GEP 模組資料

GEP 模組資料中，採用過去文獻常用於 GEP 模型的幾種技術面分析指標（張清良，2008），彙整如表二所示。技術指標的計算藉由 python 中的 Talib 套件將每日股價轉換成各種技術指標值。在股價方面選用調整後的價格，目的在於降低各股除權息造成的股價跳空。此外，本研究 GEP 模型為動態去選擇技術指標種類以及其所對應的指標天期，因此將指標的天期範圍設定為 5 至 60 日，6 種技術指標共 336 個特徵提供給 GEP 進行交易演化。

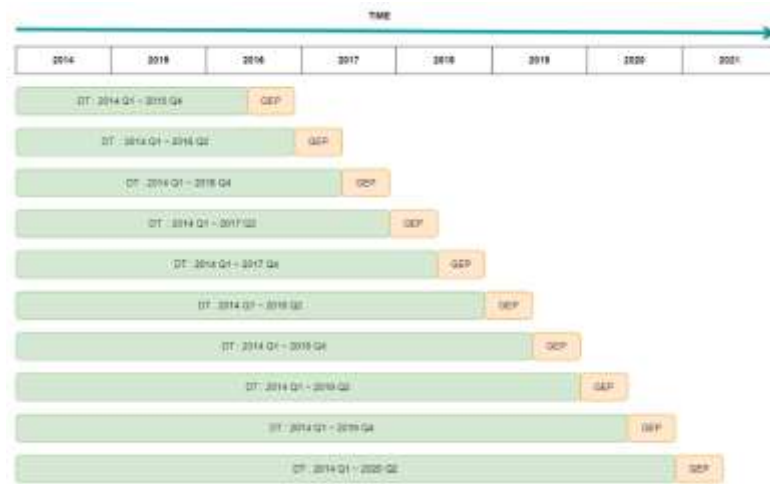
表二 技術指標表

編號	技術指標(Technical Indicator)
1	隨機指標 K
2	隨機指標 D
3	威廉指標(WMS%R)
4	相對強弱指標(RSI)
5	動量指標(MTM)
6	乖離率(BIAS)

(三) 移動窗格

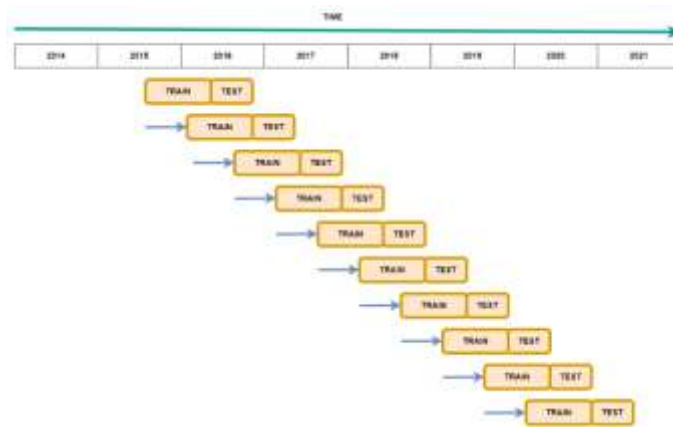
訓練模型採用李佳儒（2010）提出之移動窗格 (Walk Forward) 方式，此作法可將股價的連續時間序列資料切割成含有不同盤勢的片段資料，透過這種方式能夠評估模型於各種不同盤勢中的績效表現與穩定度。在選股與擇時模型中分別採用定錨 (Anchored) 與非定錨 (Non-Anchored) 兩種方式建模。定錨建模指每次移動窗格時，訓練集的起始點保持不變，而只移動終止點。這意味著每個新視窗都包含之前所有的歷史資料。採用定錨建模除了可以有效利用歷史資料捕捉長期趨勢和規律，還可避免當訓練集資料不斷擴大時，降低模型重新訓練的頻率；而非定錨建模則是讓每次移動窗格時，訓練集的起始點和終止點都進行相同長度的移動。這意味著每次更新訓練資料時，每個新視窗都只會包含一段固定長度的最近資料。透過，非定錨建模可以更即時地捕捉市場的短期波動訊息，以利擬定短期投資和高頻交易策略。

本研究首先藉由定錨的方式，逐期更新選股與增加訓練期資料，藉此動態更新選股策略，選股模型變化如圖六所示。



圖六 選股模型移動視窗規範圖

每一期透過選股模型所選出的股票，會作為交易標的提供給 GEP 進行建模，交易模型會透過非定錨的移動窗格方法分期訓練與測試。為平衡樣本容量、避免過度擬合與提升模型複雜度，以及考慮市場動態變化等諸多因素，本研究之 GEP 以 3 季的長度作為訓練期；此外，在評估模型於新數據上的表現，採用比訓練期相對較短的季度可以更有效捕捉短期市場變化並即時更新股池，確保模型持續關注市場中表現優秀的股票並隨時調整投資組合，鑑於此，本研究以 2 季的長度作為測試期，時間切割方式如圖七所示。



圖七 GEP 移動視窗規範圖

三、選股模組

本研究選用 RF 與 LGBM 兩種集成式決策樹增強選股模型，並使用基本面選股因子作為兩者的輸入特徵，透過從選股模型中挑選出未來上漲機率較大的潛力標的，提供給 GEP 模型進行訓練與交易。RF 擁有穩健性強的優勢，除適用於金融資料中經常發生缺失或異常情況外，通過集成多棵決策樹，可有效減少單個決策樹過度擬合之問題。此外，透過 RF 可評估各個特徵的重要性，有助決策者理解哪些因數在選股和擇時中起到了關鍵作用，非常適用於處理包含大量特徵等高維資料的選股和擇時模型；而 LGBM 則是在處理大規模資料時，通過採用梯度提升演算法，使其具備更優異的效率與精確度。除此之外，占用之運算資源亦較低，適合處理高頻交易與擁有海量金融資料的分析。另一方面，LGBM 可以直接對類別屬性特徵進行處理而無需進行獨熱編碼 (One-Hot Encoding)，簡化了資料預處理流程。

關於決策樹貼標方式，本研究以計算報酬率作為依據，將未來兩季 (T+2) 漲幅大於 0 的股票標記為上漲，反之為下跌。此外，在選股方面若只選擇單一商品作為交易標的，可能承受較大的交易風險。因此，本研究以兩季為單位更新一次選股池，並在每次選股的預測結果中產生出包含五檔股票的投資組合作為交易標的。至於在決策樹模型參數設計上，並無特別調整模型參數，僅採用兩類模型之預設參數值。

四、GEP 模組

本研究使用基因表達規劃法訓練一套交易策略模型，可稱為智慧型交易策略產生機器人。在基因編碼部分採取 Ferreira (2001) 所提出的多基因編碼方式，以不同的表達樹相互作用，形成一個能解決複雜問題的多基因表達樹。本研究基因包含交易訊號、資金管理與停損停利三種基因。以下分別介紹各個基因所包含的功能以及相關的操作。

(一) 基因一、二：交易訊號與資金管理

此基因功能在於產生進出場點（交易訊號）與買賣張數（資金管理），在多種技術指標、天期與各指標閾值比率下找出交易期間內最合適的策略，並搭配隨機數值常數產生每種交易動作最佳的買賣張數。

編碼方式由頭部 (Head)、尾部 (Tail)、DC-specific 與 RNCs 陣列所構成，在基因的頭部中除了根節點必須是屬性節點（技術指標）外，其他的節點可以

是屬性節點（技術指標）也可以是終端節點（交易訊號），這是為了避免造成此染色體變成一條無效的規則。而在尾部中節點只能是終端節點（交易訊號），則是為了規範交易規則樹不可無限制的生長。

屬性節點由各種技術指標所組成，每個屬性節點的分支都為 3，分支可以看作是交易規則的延展。換句話說一個屬性節點可以代表著三種可能的交易規則，若其下分支的子節點也是屬性節點則代表再添加一層指標規則。而子節點若是終端節點則代表最終的交易動作，可能為買入(1)、不動作(0)與賣出(-1)三種訊號。交易訊號基因的屬性與終端節點集如表三。

表三 交易訊號屬性與終端節點表

節點類別	節點名稱	代表符號	分支
屬性節點	隨機指標 K	K	3
	隨機指標 D	D	3
	相對強弱指標	RSI	3
	乖離率	BIAS	3
	動量指標	MOM	3
	威廉指標	WTLLR	3
終端節點	買入	1	0
	不動作	0	0
	賣出	-1	0

交易訊號基因中包含一個 DC-specific，內含四個部分，分別代表指標天期、指標最小閾值、指標最大閾值與買賣張數，相關值域範圍如表四所示。天期指基因中每個屬性節點（技術指標）所對應的隨機天期，值域介於在 5 至 60 日之間，例如技術指標可以是五日的 RSI、十日的 RSI 或是六十日的 RSI。最小與最大指標閾值則作為每個屬性節點(技術指標)的規則範圍，也就是技術指標的交易規則判斷；而買賣張數為每個終端節點(交易動作)所對應的部位大小。

表四 交易訊號基因 DC 值域表

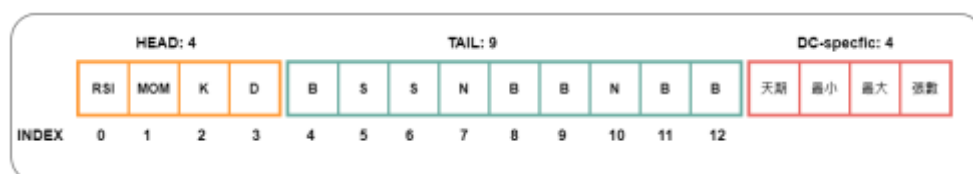
DC-specific 名稱	對應節點名稱	值域	DC-specific 名稱	對應節點名稱	值域
技術指標天期	屬性節點	[5,60]	指標閾值最大值	屬性節點	[0,1]
指標閾值最小值	屬性節點	[0,1]	買賣張數大小	終端節點	[1,5]

在交易訊號基因長度參數如表五所示，本研究於設計上，為了使 DC-specific 都能找到其所對應的值，將天期、最小與最大閾值三個 DC-specific 的陣列長度對照頭部的長度，也就是一個基因中所包含技術指標的數量；而買賣張數的 DC-specific 長度計算方法則是尾部加上頭部減去根節點，這是由於 GEP 的特性，最簡短的交易規則可能只有一個技術指標配上三種交易訊號。

表五 交易訊號基因參數表

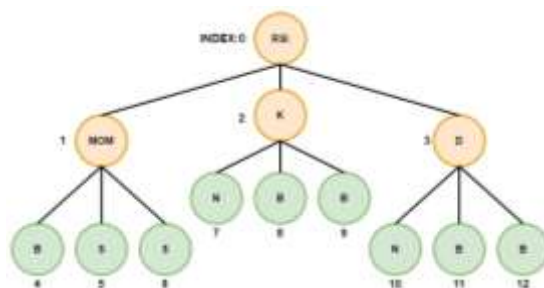
參數名稱	長度	值域	參數名稱	長度	值域
頭部	6	技術指標	DC_買賣張數	18	[0, 35]
尾部	13	交易訊號	Rnc_天期	12	[5, 60]
DC_天期	6	[0, 11]	Rnc_最小閾值	12	[0, 1]
DC_最小閾值	6	[0, 11]	Rnc_最大閾值	12	[0, 1]
DC_最大閾值	6	[0, 11]	Rnc_買賣張數	36	[1, 5]

GEP 模型在演化時，會由線性染色體進行染色體再製，而在評估適應函數時則會將線性染色體透過 K 表達式將染色體轉為規則樹。以一個頭部長度為 4 的交易訊號基因作為交易規則之判斷為例，尾部經過計算能得到長度為 9 之基因尾部，其產生之線性染色體如圖八所示。



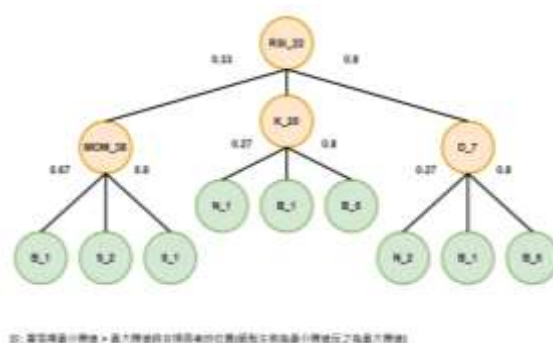
圖八 交易訊號線性染色體表示圖

接下來，將此線性染色體頭部與尾部中的節點，依序按照上至下左至右的順序，將節點填入並轉換為表達樹。此流程是先填入一個節點後判斷此節點的分支數量並向下延展子節點，若當前節點分支數量為 0 則沒有向下分支，若分支數大於 0 則重複此步驟。可以從圖九的線性染色體節點索引位置去對照圖三至六表達樹節點。



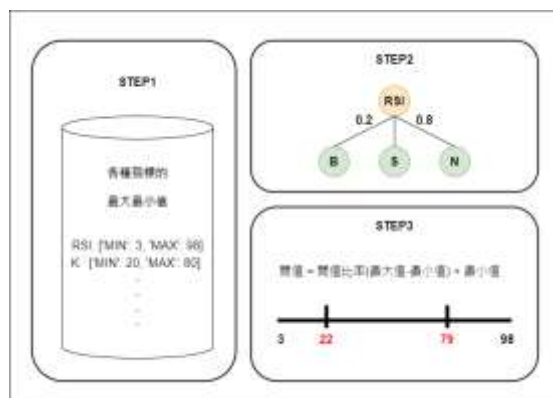
圖九 交易訊號基因表達樹示意圖

接著第二步從表達樹中依序由上至下左至右讀取節點，如果是屬性節點會得到三種 DC 值，分別是技術指標的天期、技術指標最小閾值比率與技術指標最大閾值比率，若是終端節點的話只會得到買賣張數的 DC，每個節點所得到的 DC 值都代表著一個索引位置。這些索引位置能夠從其專屬的 RNC 陣列中找到相對應的資訊。交易訊號基因之 DC-specific 與 RNC 陣列如圖十所示。



圖十二 交易訊號基因最終表達樹示意圖

另外值得一提的是，關於技術指標規則在產生之前，都會先從訓練期資料中讀取各個指標在此區間內的最大與最小值並存放。每個指標亦有獨自的一個區間範圍且不互相干擾。以圖十二之根節點 **RSI_22** 為例，如果在訓練期區間內 22 日的 **RSI** 值最小是 3 而最大是 98 的話，這時再與剛剛所得到的指標最小閾值與最大閾值比率做計算，即可產生交易規則，流程如圖十三所示。



圖十三 指標規則產生示意圖

以圖十三為例，當 22 日的 **RSI** 值小於 22 時規則往左，介於 22 至 79 時往中，而大於 79 則往右。當得到當前技術指標的左、中、右三種規則方向後，重複此步驟直到當前節點分支數為 0（即終端節點）時則停止規則判斷，此基因最終總共分別產出 9 種規則：

規則 1：22 日的 **RSI** 值處於低值 且 38 日 **MTM** 處於低值時 買一張。

規則 2：22 日的 **RSI** 值處於低值 且 38 日 **MTM** 處於中值時 賣兩張。

規則 3：22 日的 **RSI** 值處於低值 且 38 日 **MTM** 處於高值時 賣一張。

規則 4：22 日的 **RSI** 值處於中值 且 20 日 **K** 處於低值時 不動作。

規則 5：22 日的 RSI 值處於中值 且 20 日 K 處於中值時 買一張。

規則 6：22 日的 RSI 值處於中值 且 20 日 K 處於高值時 買五張。

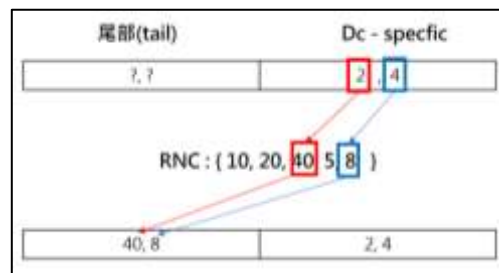
規則 7：22 日的 RSI 值處於高值 且 7 日 D 處於低值時 不動作。

規則 8：22 日的 RSI 值處於高值 且 7 日 D 處於中值時 買一張。

規則 9：22 日的 RSI 值處於高值 且 7 日 D 處於高值時 買五張。

(二) 基因三：風險管理

當手上持有部位時，會藉由基因三作為風險管理的一環，由手上的部位與商品目前的收盤價計算當前的未實現損益，當損益達到停損停利點時便進行全部出場的動作。此基因編碼頭部及尾部長度分別為 0 與 2，基因僅由尾部、DC-specific 與 RNC 陣列所組成，尾部中的兩終端節點（即停損點與停利點）皆會以隨機數值常數符號「？」表示，再由 DC-specific 找出 RNC 陣列中所對應的停損停利點，透過所設定的值域建立交易策略最佳的停損停利機制。以圖十四為例，假若 DC-specific 長度為 2，而 RNC 長度為 4 且值域介於 0 至 50 之間，將產生 5 個隨機數值。當手上持有部位時，若交易策略的報酬率獲利達到 40% 時以停利方式出場，或是當交易策略損失達到 8% 時以停損方式出場。



圖十四 基因三之停損停利染色體編碼

(三) 交易成本(trade cost)

在 GEP 交易模型中進行訓練與測試時為了模擬實際交易結果，在每次股票交易買賣時都會扣除手續費以及交易稅成本。當在進行買賣時都會收取一筆券商的手續費，而只有在賣出時會收取 0.3% 證卷交易稅，在本研究中將交易手續費設為 0.1425%，每一次交易後須扣除的交易成本。計算方式如公式(1)：

券商手續費：0.1425%

證卷交易稅：0.3%

買進總額：買進價格 * 股數(1000) * 張數

賣出總額：賣出價格 * 股數(1000) * 張數

$$\text{交易成本} = (\text{買進總額} + \text{賣出總額}) * \text{手續費} + \text{賣出總額} * \text{交易稅} \cdots \cdots (1)$$

(四) 適應函數(fitness function)

本研究為了使交易策略獲利最大化，採用絕對報酬率 (ROI) 作為族群內每一個染色體的適應值。絕對報酬率表示在一段期間內策略所獲得的總淨利和初始資金的比值。其中總淨利為每次交易時所獲得的損益加總，每筆損益都有扣除每次所產生的交易成本。適應函數的計算方式如公式(2)：

$$TP = \text{每期交易總損益}$$

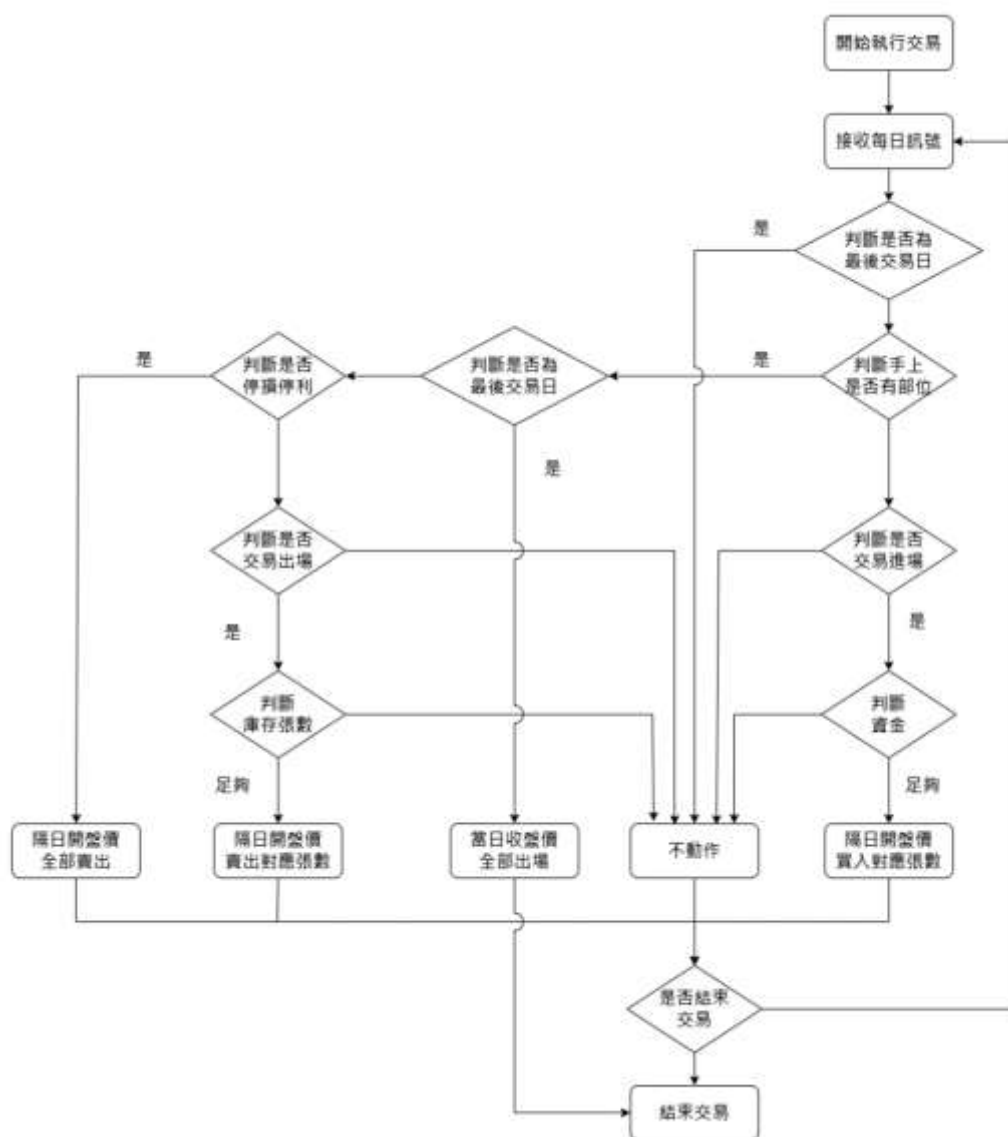
$$M = \text{每期投入本金}$$

$$F = TP / M \cdots \cdots (2)$$

(五) 交易方式

在本研究中，適應值是由設定的交易區間內，經過每一次的交易產生的損益所計算而成。在區間內逐日輸入當天的指標值並透過規則樹轉換成訊號，在進行交易前會先判斷是否為最後交易日，若是則不進行動作反之則進入訊號判斷；當手上沒有部位時判斷交易訊號，而當出現買入訊號時透過資金管理基因找出買入張數，並判斷資金是否充足，若手上資金足夠則買入反之則不動作；當手上持有部位時會藉由風險基因判斷目前是否該執行停損停利機制，若手上目前部位與當日收盤價的預估損益達到門檻時則全部出場；而當手上有部位不需要停損停利但出現賣出訊號時，會判斷手上持有的股票張數是否多於規則樹產生的賣出張數，若大於則賣出對應張數反之則不動作；若未達所設定之最終交易日則重複執行上述步驟，經過以上流程則完成一次整個的交易流程。

然而，由於訓練期目的在於找出一有效之交易策略，並無強制在最後交易日出場所有部位；但測試期為了能夠評估每一期的交易績效及分配資金，因此，本研究於交易流程中加入最後交易日所有部位強制出場的條件，如圖十五所示。



圖十五 測試期交易流程圖

五、實驗設計

本研究在於探討 GEP 使用動態選股的方式後其績效變化，共分成三個實驗進行比較分析，實驗說明如下：

(一) 實驗一(純 GEP 交易策略)

使用 GEP 交易策略模型對選股模型第一期所選出的標的進行交易（時間區間為 2016/05/16 至 2021/05/15），評估在不使用動態更新選股的情況下，GEP 交易模型其績效會如何改變。

(二) 實驗二(動態選股 GEP 交易策略)

使用動態選股每兩季更新一次投資組合，並搭配 GEP 交易策略，評估使用動態投資組合後交易績效是否有提升。

(三) 實驗三(綜合分析比較)

使用各種選股因子於「不同」集成決策樹模型，探討其所產生的股票標的在買入持有的方式下，與純 GEP 和動態選股 GEP 兩種方法的交易表現之優劣。

六、績效評估模組

本研究評估分為選股模型以及 GEP 擇時模型兩個部分，在選股模型中以分類的結果評估模型的表現，而在 GEP 模型中以交易績效指標來分析各期與總交易的策略結果。

(一) 選股模型評估

本研究選股模型以分類的方式預測未來股價之漲跌，因此在模型表現方面使用混淆矩陣作為評估指標，模型預測結果之指標包含準確率 (Accuracy)、精確率 (Precision)、召回率 (Recall) 及 F1 值。

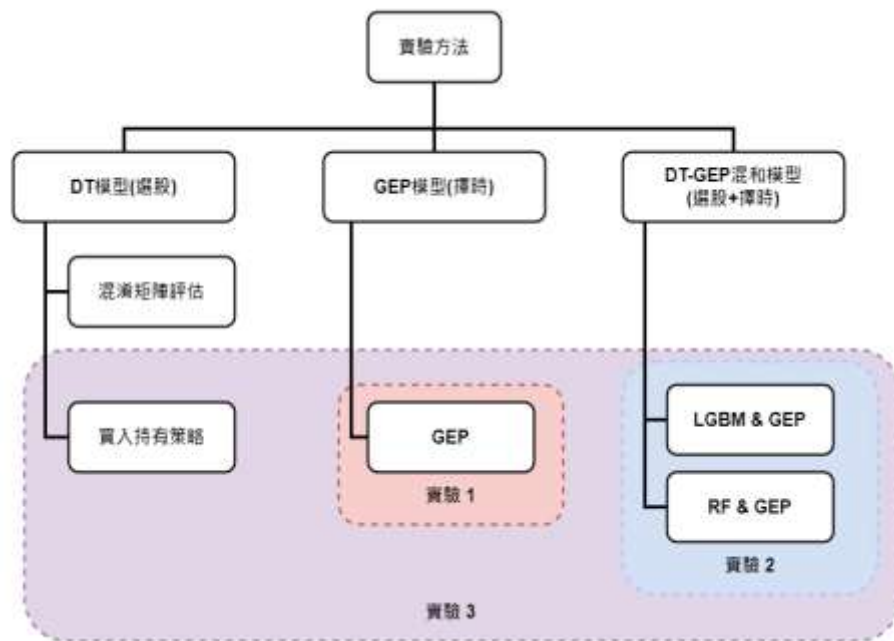
(二) 交易策略評估

在交易績效評估之範圍將考慮到獲利、風險、交易三個分析層面，進而驗證本研究投資策略之有效性。而三個分析層面包含淨利 (Net Profit)、絕對報酬率 (Return on Investment, ROI)、獲利因子 (Profit Factor)、夏普率 (Sharp Ratio)、最大交易回落 (Max Drawdown, MDD)、風險報酬比 (Reward to Risk Ratio)、

年化報酬率 (Annualized Return)、勝率 (Winning Rate)、盈虧比 (Profit and Loss Ratio) 與凱利值 (Kelly) 等個別詳細績效指標。

肆·實驗結果與分析

根據前述研究方法，本研究將訓練並測試選股與交易兩種模型，並依序介紹說明所使用之硬軟體設備、選股模型表現與三種實驗結果。實驗工作分配如圖十六所示。本研究將實驗分為選股、擇時與混合應用三個部分，在選股方面主要評估兩種決策樹模型於測試期表現，並以混淆矩陣評估兩者模型之優劣；在擇時部分探討不更新選股的 GEP 模型，在只使用技術指標的情況下是否依舊能在股市中具有穩定的獲利能力；而混合應用為同時使用選股與擇時兩種方法的交易模型。



圖十六 實驗工作圖

一、實驗環境與工具

本研究根據研究所需，遴選適合的軟體及硬體來架設實驗環境，各軟硬體設備規格如表六所示。

表六 實驗硬體設備及軟體環境表

硬體設備 / 軟體環境	型號 / 版本
處理器	Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz
記憶體	16GB
硬碟	1TB
顯示卡	GeForce GTX1660
作業軟體	Microsoft Windows10
程式開發工具	Spyder
程式語言	Python 3.8

二、選股模型評估

本節探討 RF 與 LGBM 兩種決策樹模型於各個時期的預測表現，起始時間從 2016/5/15 至 2020/11/14 共 10 期，透過移動窗格方法預測 T+2 期（即六個月後）的漲跌。使用資料為每一期中所包含之財報，特徵包含規模、品質、價值、動能、高股息與低波動等因子作為輸入特徵。表七與八為兩種選股模型之參數設定。

表七 RF 參數設定表

參數名稱	參數值	說明
n_estimators	100	森林中樹木的數量
max_depth	None	每顆樹的最大深度
min_samples_leaf	1	葉子節點最少樣本數
min_samples_split	2	內部節點再劃分所需最小樣本數
max_features	auto	每棵樹最大特徵數量
criterion	gini	分割特徵的方法

表八 LGBM 參數設定表

參數名稱	參數值	說明
num_iterations	100	總共迭代次數
learning_rate	0.1	學習速率
boosting	gbdt	Boosting 種類

max_depth	-1(無限制)	樹的最大深度
min_data_in_leaf	20	一個子葉中最少數據
max_bin	255	將特徵值放入桶中的最大 bins 數

表九與十分別為 RF 與 LGBM 選股模型實驗結果，透過定錨的移動窗格方法在每一期產生預測結果後，以混淆矩陣評估各期的模型表現。在沒有特別優化模型參數之情況下，以平均值來看 LGBM 的選股模型在 4 種指標的表現均略高於 RF。單以準確率來觀察兩種模型的表現能發現，在預測整體漲跌的部分兩者準確率皆在五至六成左右。然而在本研究中，選股目的在於找出未來具有上漲可能之股票，故精確率是較為重要的一項指標。從兩種選股模型的各期表現來看，在不同時期中精確率表現相差無幾，但能發現在第四、五、八期時，兩種模型的精確率皆低於五成，原因可能為有重大外來因素影響整體股票市場的趨勢，這也代表兩種模型皆無法在這類時期下有較佳的預測表現。

表九 RF 模型指標評估表

各期 \ 指標	準確率 (Accuracy)	精確率 (Precision)	召回率 (Recall)	F1 Score
2016/05/15	0.511	0.709	0.418	0.526
2016/11/14	0.515	0.780	0.503	0.612
2017/05/15	0.561	0.687	0.559	0.616
2017/11/14	0.557	0.469	0.473	0.471
2018/05/15	0.646	0.329	0.448	0.379
2018/11/14	0.501	0.846	0.460	0.596
2019/05/15	0.489	0.704	0.339	0.457
2019/11/14	0.602	0.398	0.459	0.426
2020/05/15	0.489	0.810	0.472	0.596
2020/11/14	0.564	0.599	0.539	0.568
平均	0.544	0.633	0.467	0.525

表十 LGBM 模型指標評估表

各期 \ 指標	準確率 (Accuracy)	精確率 (Precision)	召回率 (Recall)	F1 Score
2016/05/15	0.582	0.704	0.613	0.656
2016/11/14	0.657	0.814	0.710	0.759
2017/05/15	0.582	0.661	0.693	0.677
2017/11/14	0.581	0.499	0.698	0.582
2018/05/15	0.654	0.365	0.591	0.452
2018/11/14	0.579	0.842	0.583	0.689
2019/05/15	0.515	0.763	0.345	0.475
2019/11/14	0.558	0.372	0.545	0.443
2020/05/15	0.683	0.819	0.775	0.796
2020/11/14	0.608	0.611	0.718	0.660
平均	0.600	0.645	0.627	0.619

三、純GEP交易模型

本節探討在不使用動態選股的情況下，只使用技術指標作為買賣參考依據的 GEP 模型，是否有在各種趨勢中獲利的能力。GEP 交易模型之參數參考至呂冠融 (2018) 所設計 GEP 模型，僅在單點重組以及雙點重組提高比率，目的在於使模型加快收斂，交易模型之演化機制與基因參數設定如表十一所示。

表十一 GEP 交易模型演化機制與基因參數表

項目	參數名稱	參數值
演化機制	迭代次數	50
	族群大小	50
	突變率	0.2
	反轉率	0.1
	插入轉換率(IS)	0.15
	根插入轉換(RIS)率	0.15
	單點重組率	0.3
	雙點重組率	0.3
	DC 突變率	0.2
	DC 反轉率	0.2
	DC 轉置率	0.2
	RNC 突變率	0.15
交易訊號 與 資金管理	頭部長度	6
	最大分支	3
	尾部長度	13
	屬性節點	K,D, RSI, BIAS, WILLR, MOM
	終端節點	買入(1)，不動作(0)，賣出(-1)
	Dc1 長度(指標天期)	6
	Dc2 長度(指標最小閾值)	6
	Dc3 長度(指標最大閾值)	6
	Dc4 長度(買賣張數)	18
	RNC1 長度(指標天期)	12
	RNC1 值域	[5, 60]
	RNC2 長度(指標最小閾值)	12
風險管理	RNC2 值域	[0, 1]
	RNC3 長度(指標最大閾值)	12
	RNC3 值域	[0, 1]
	RNC4 長度(買賣張數)	36
	RNC4 值域	[1, 5]
	頭部長度	0
	最大分支	0
	尾部長度	2
	屬性節點	-
	終端節點	?
	Dc 長度(停損、停利)	2
	RNC 長度(停損、停利)	4
	RNC4 值域	[1, 50]

(一) 實驗說明

本節實驗使用選股模組所提供之第一期投資組合作為交易標的，投資組合中包含五檔不同的股票，如表十二所示，每個標的使用之交易資金皆為平均分配。在本研究中，GEP 交易模型初始資金設為 1000 萬，分配的方式是在每一期開始前將資金平均分配。如第一期每檔股票將分得 200 萬初始本金，在第一期結束後五檔股票總淨利 200 萬，本金更新為 1200 萬，並平均分配給下一期投資組合進行交易；第二期每檔股票分得 240 萬初始本金，以此類推。

表十二 實驗一股票代號表

決策樹模型	標的一股票代號	標的二股票代號	標的三股票代號	標的四股票代號	標的五股票代號
RF	6414	6409	4426	8446	2228
LGBM	6414	6409	9910	3205	2231

(二) 實驗結果

此實驗中 GEP 交易模型皆執行十次並取平均值用以分析比較績效結果，以獲利、風險、總交易與各期交易評估此實驗模型的表現。

1. 獲利面

在獲利方面單只使用 GEP 交易模型進行買賣的方式，其獲利能力表現並不出色，兩種投組在多次實驗下皆產生虧損，在總淨利、絕對報酬率、年化報酬率與獲利因子等方面之平均表現均不如預期，結果如表十三所示。

表十三 無選股投組總績效（獲利）報表

決策樹模型 \ 指標	獲利分析			
	總淨利	絕對報酬率	年化報酬	獲利因子
RF	-937412	-9.37%	-1.98%	0.671
LGBM	-870581	-8.71%	-1.84%	0.677

2. 風險面

在不選股僅依靠技術指標所設計的 GEP 交易模型，其產生的 MDD 是非常大的，且在十次平均的夏普值中兩者皆為負值，代表以此模型所產生的交易策略其風險是高過於報酬的，結果如表十四所示。

表十四 無選股投組總績效（風險）報表

決策樹模型 \ 指標	風險分析			
	夏普值	MDD	MDD(%)	風報比
RF	-0.249	1986756	511.67%	-0.453
LGBM	-0.256	1849403	470.83%	-0.432

3. 總交易面

根據表十五顯示，兩種投組的勝率平均約在五成上下，在盈虧比方面兩者皆小於 1，代表在平均每次交易皆為虧損大過於報酬，而從凱利值也能看出在十次的結果下，兩種投組皆不適合投入資金。

表十五 無選股投組總績效（交易）報表

決策樹模型 \ 指標	總交易分析			
	盈虧比	勝率	凱利值	交易次數
RF	0.730	52.00%	-0.16	223.6
LGBM	0.512	53.65%	-0.46	154.3

(三) 各期交易

儘管在上述總交易的各項績效指標來看，只使用 GEP 模型的交易方式不能獲取正報酬，但從表十六之各期的交易績效表現能發現，GEP 交易模型並非在每一期皆處於虧損狀態，像第 6、7、9 期兩種投資組合皆有不錯的正報酬表現。

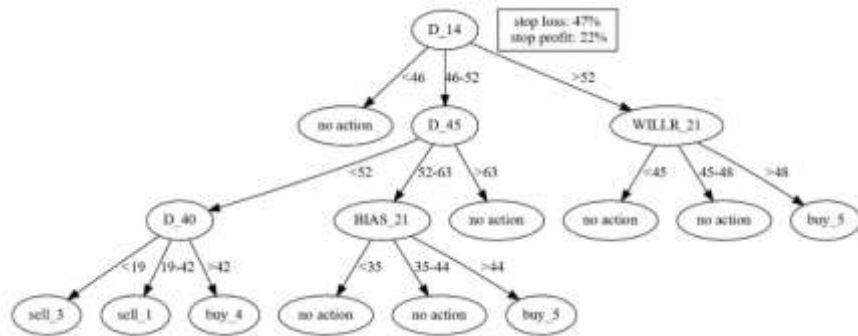
表十六 實驗一各期表現彙整分析表

期 間	各期交易績效							
	年化報酬率		平均報酬率		勝率		MDD(%)	
	RF	LGBM	RF	LGBM	RF	LGBM	RF	LGBM
1	0.31%	-5.83%	0.43%	-6.12%	32%	83%	109%	201%
2	-7.85%	4.51%	-4.01%	2.51%	50%	46%	953%	39%
3	-0.38%	2.86%	-0.30%	2.64%	56%	39%	168%	66%
4	-0.87%	6.91%	-0.53%	4.82%	66%	58%	309%	62%
5	-20.96%	-19.94%	-11.88%	-13.11%	45%	20%	2471%	1363%
6	17.96%	5.38%	12.81%	3.82%	88%	62%	6%	62%
7	7.22%	3.46%	4.09%	2.93%	65%	41%	27%	139%
8	-14.47%	-20.41%	-9.12%	-12.68%	12%	10%	1346%	1569%
9	12.19%	11.43%	6.45%	11.85%	64%	89%	17%	5%
10	-5.65%	0.81%	-2.90%	0.50%	23%	61%	367%	410%

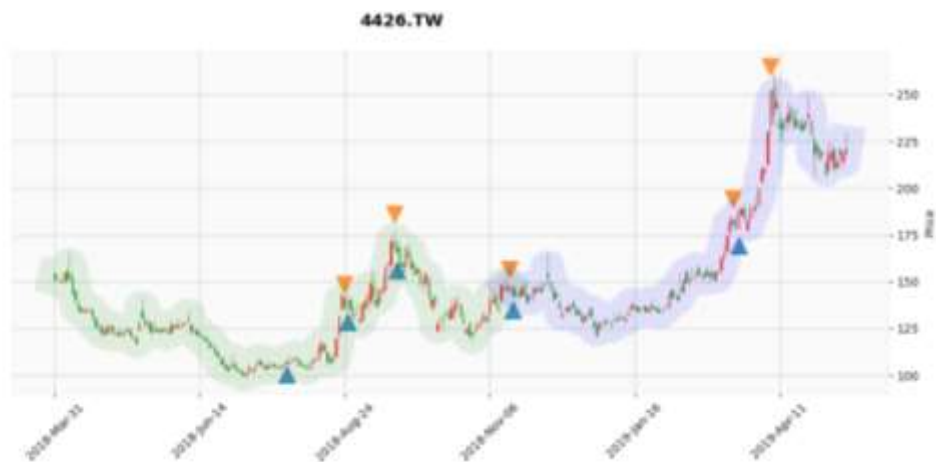
(四) 實際交易情況及演化結果

以各期之績效最佳策略來看，表現最佳為 RF 投組中第六期的實驗結果，當期年化報酬率為 17.96%，投資組合商品分別為樺漢 (6414.TW)、旭隼 (6409.TW)、利勤 (4426.TW)、華研 (8446.TW) 與劍麟 (2228.TW) 五檔交易標的。其中，表現最佳的為利勤 (4426.TW)，其最佳的適應函數（絕對報酬率）在訓練期中能達到 17% 左右，以訓練其中最佳適應函數所產生的表達樹會如圖十七所示，透過此表達樹分別在訓練期與測試期間找出訊號並執行交易，實際的交易如圖十八所示（藍色箭頭為進場，橘色箭頭為出場，綠色區塊為訓練期間，紫色區塊為測試期間），從實際交易能發現在測試期時能有效地找到最

佳的進出場點，訓練與測試期走勢相似，從訓練時所得到的交易模型能夠實戰應用於測試期中。

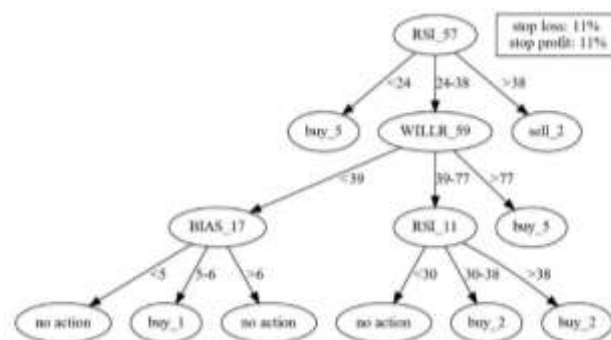


圖十七 利勤 (4426.TW) 第六期最佳基因表達樹

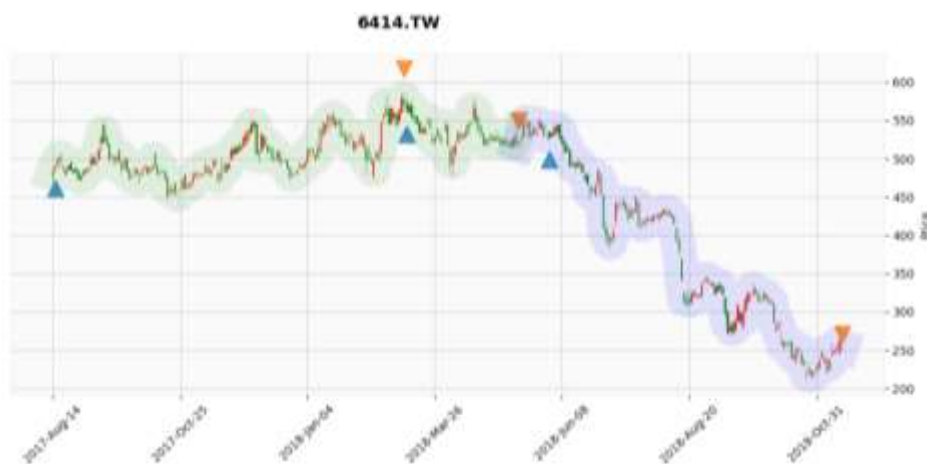


圖十八 利勤 (4426.TW) 第六期實際交易示意圖

以各期之績效最差策略來看，表現最差為 RF 投組中第五期的實驗結果，當期年化報酬率為-20.96%，投資組合商品分別為樺漢 (6414.TW)、旭隼 (6409.TW)、利勤 (4426.TW)、華研 (8446.TW) 與劍麟 (2228.TW) 五檔交易標的。其中，表現最差的為樺漢 (6414.TW)，其最佳的適應函數(絕對報酬率)在訓練期中能達到 11%左右，以訓練其中最佳適應函數所產生的表達樹會如圖十九所示，透過此表達樹分別在訓練期與測試期間找出訊號並執行交易，實際的交易如圖二十所示，從實際交易能發現在訓練期時走勢處於盤整，雖然在訓練期時模型能從中找出獲利 10%以上的交易策略，但此策略用於測試期則明顯找不出合適的交易進出場點，直至最終交易日強制出場機制觸發。



圖十九 樺漢 (6414.TW) 第五期最佳基因表達樹



圖二十 樺漢 (6414.TW) 第五期實際交易示意圖

四、決策樹與GEP混合模型

(一) 實驗說明

本節實驗使用選股模型結合 GEP 進行訓練與交易，以動態選股的方式更新每一期 GEP 所需交易之標的物，以每兩季更新一次的頻率動態調整投組中的股票，兩種決策樹模型於各期中所預測之標的物如表十七所示。

表十七 實驗二各期選股表

代號 各期	決策樹 模型	標的一 股票代號	標的二 股票代號	標的三 股票代號	標的四 股票代號	標的五 股票代號
2016/5/15	RF	6414	6409	4426	8446	2228
2016/11/14		2330	3450	3324	6187	1590
2017/5/15		6409	6505	6411	5536	8436
2017/11/14		1301	1909	5264	4528	4912
2018/5/15		8436	3406	2492	2330	2104

表十七 實驗二各期選股表（續）

代號 各期	決策樹 模型	標的一 股票代號	標的二 股票代號	標的三 股票代號	標的四 股票代號	標的五 股票代號
2018/11/14	RF	2492	2327	6409	3034	6274
2019/5/15		3008	5274	8499	5347	3293
2019/11/14		3406	6409	6176	2379	5274
2020/5/15		3008	6538	3227	6415	2330
2020/11/14		3661	2330	2327	1256	2458
2016/5/15	LGBM	6414	6409	9910	3205	2231
2016/11/14		2330	9910	2395	6230	6415
2017/5/15		6415	5274	6505	2723	2330
2017/11/14		5274	2360	2723	8436	4943
2018/5/15		4807	3227	3515	8436	8437
2018/11/14		8436	8489	3034	6409	5274
2019/5/15		3008	2379	9910	5274	6488
2019/11/14		6462	5274	8437	2379	3293
2020/5/15		3661	2379	3008	9910	2330
2020/11/14		3661	2454	2379	2330	3034

(二) 實驗結果

此實驗中 GEP 交易模型皆執行十次並取平均值用以分析比較績效結果，以獲利、風險、總交易與各期交易評估此實驗模型的表現。

1. 獲利面

在獲利方面，不管在總淨利、絕對報酬率、年化報酬率下，兩種投組的交易績效皆有不錯表現。以平均來看，兩者年化報酬皆高於 5% 以上，在獲利因子方面兩者平均皆高於 15，代表在每輸一元下能賺回 15 塊以上，結果如表十八所示。

表十八 動態選股投組總績效（獲利）報表

決策樹模型	指標	獲利分析			
		總淨利	絕對報酬率	年化報酬	獲利因子
RF		5699605	57.00%	9.16%	25.939
LGBM		2978806	29.79%	5.25%	19.466

2. 風險面

從 MDD 與風報比兩者表現來看，儘管 MDD 的表現在使用動態選股後的 GEP 交易模型，其風險方面表現有所降低依舊不甚理想，但由於交易策略所獲得的報酬表現使得風報比的數值並未因風險而有所降低。在夏普值方面，雖兩種投組產生的值皆為正值 (0.495、0.596)，但交易策略所產生的平均報酬波動太大，以至夏普值的表現不盡理想，結果如表十九所示。

表十九 動態選股投組總績效（風險）報表

決策樹模型 \ 指標	風險分析			
	夏普值	MDD	MDD(%)	風報比
RF	0.596	423158	125.41%	16.730
LGBM	0.495	689013	501.56%	10.212

3. 總交易面

根據表二十顯示，在勝率方面，RF 投組的平均表現略勝於 LGBM 投組，盈虧比方面兩者則相差不大，皆大於 5，代表每次交易下，贏時所獲得的金額是高過於輸時所造成的虧損，而從凱利值比較勝率與盈虧比兩者間的關係發現，RF 投組在十次的結果下是優於 LGBM 投組的。

表二十 動態選股投組總績效（交易）報表

決策樹模型 \ 指標	總交易分析			
	盈虧比	勝率	凱利值	交易次數
RF	5.082	66.22%	0.52	238.7
LGBM	5.312	54.80%	0.19	111.3

(三) 各期交易

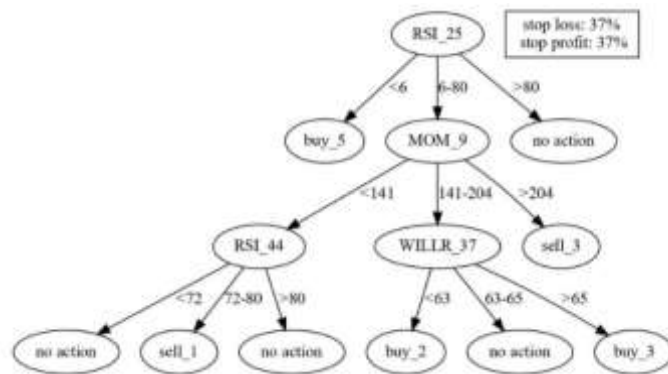
從表二十一可發現，動態選股 GEP 模型之 2、3、6、7、9、10 期，兩種投組在年化報酬以及平均報酬皆產生正報酬；另一方面，以表現較差的 1、4、5 期來看，雖無法優於其他期間的績效，但在同一期中儘管其中一種投組表現不甚理想，另一投組卻依然能獲取正報酬，代表透過選股模型是有機會能夠規避虧損風險。

表二十一 實驗二各期表現彙整分析表

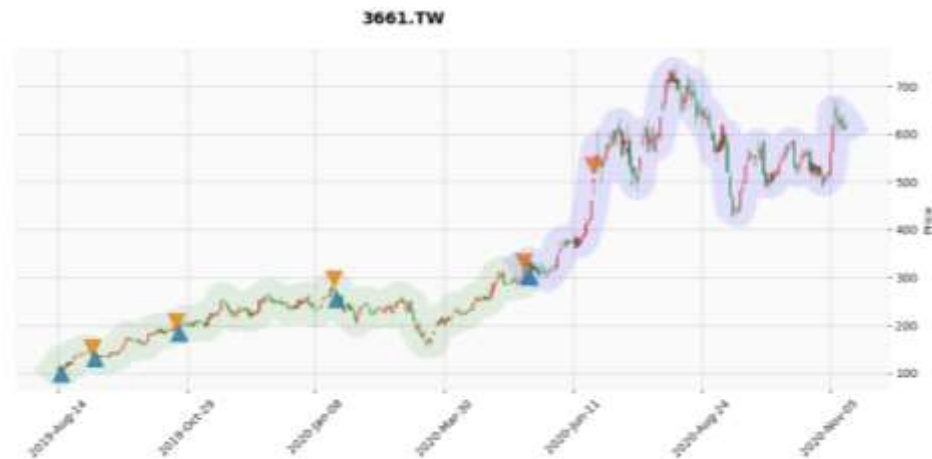
期間	各期交易績效							
	年化報酬率		平均報酬率		勝率		MDD(%)	
	RF	LGBM	RF	LGBM	RF	LGBM	RF	LGBM
1	0.02%	-4.50%	0.06%	-4.15%	34%	46%	125%	324%
2	19.48%	2.64%	8.91%	1.94%	82%	55%	0%	0%
3	9.96%	7.01%	6.04%	7.00%	79%	93%	1%	0%
4	-4.59%	4.26%	-2.20%	3.10%	35%	48%	281%	4.63%
5	10.96%	-8.47%	5.87%	-5.69%	54%	21%	81%	1370%
6	12.14%	10.59%	7.06%	9.85%	81%	73%	18%	17%
7	11.34%	9.16%	9.17%	10.98%	77%	100%	9%	0%
8	10.01%	2.44%	5.63%	1.21%	64%	33%	12%	57%
9	14.96%	25.28%	11.70%	35.93%	67%	99%	9%	1%
10	15.57%	12.45%	8.87%	16.07%	74%	58%	49%	11%

(四) 實際交易情況及演化結果

以各期之績效最佳的策略來看，表現最佳為 LGBM 投組中第九期的實驗結果，當期年化報酬率為 25.28%，投資組合商品分別為台積電 (2330.TW)、世芯-KY (3661.TW)、瑞昱 (2379.TW)、大立光 (3008.TW)、豐泰 (9910.TW) 五檔交易標的，其中表現最佳的為世芯-KY (3661.TW)，其最佳的適應函數(絕對報酬率)在訓練期中能達到 50% 左右，以訓練其中最佳適應函數所產生的表達樹會如圖二十一所示，透過此表達樹分別在訓練期與測試期間找出訊號並執行交易，實際的交易如圖二十二所示(藍色箭頭為進場，橘色箭頭為出場，綠色區塊為訓練期間，紫色區塊為測試期間)，從實際交易示意圖能看出在測試期前段能找到不錯的進出場點，但當中後期股市波動後則無法產生交易訊號。



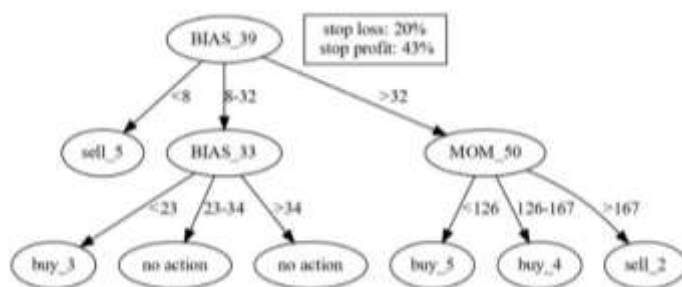
圖二十一 世芯-KY (3661.TW) 第九期最佳基因表達樹



圖二十二 世芯-KY (3661.TW) 第九期實際交易示意圖

以各期中表現最差的策略來看，表現最差為 LGBM 投組中第五期的實驗結果，當期年化報酬率為-4.59%，投資組合商品分別為日成-KY (4807.TW)、

原相 (3227.TW)、華擎 (3515.TW)、大江 (8436.TWO)、大地-KY (8437.TWO) 五檔交易標的，其中表現最差的為原相 (3227.TW)，其最佳的適應函數（絕對報酬率）在訓練期中能達到 20% 左右，以訓練其中最佳適應函數所產生的表達樹會如圖二十三所示，透過此表達樹分別在訓練期與測試期間找出訊號並執行交易，實際的交易如圖二十四所示，從實際交易示意圖能看出在測試期中期時找到進場點，但在交易結束前都無法產生出場訊號，以至於達到最終交易日的強制出場。



圖二十三 原相 (3227.TW) 第五期最佳基因表達樹

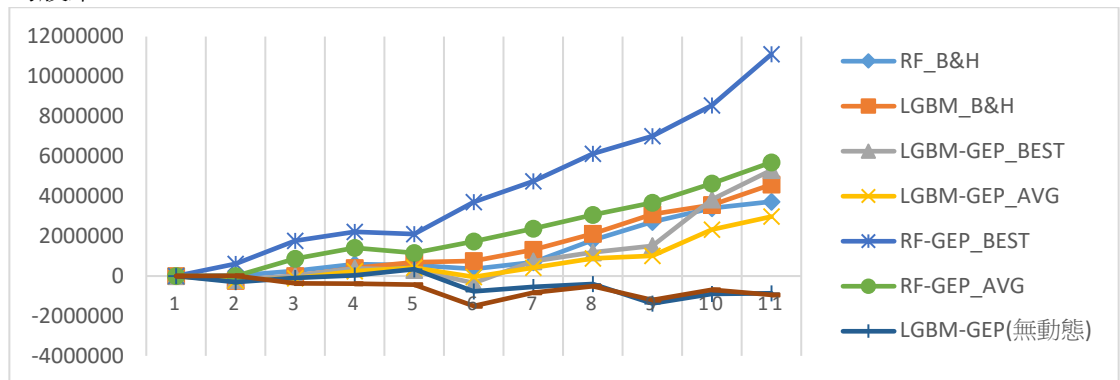


圖二十四 原相 (3227.TW) 第五期實際交易示意圖

五、綜合分析績效比較

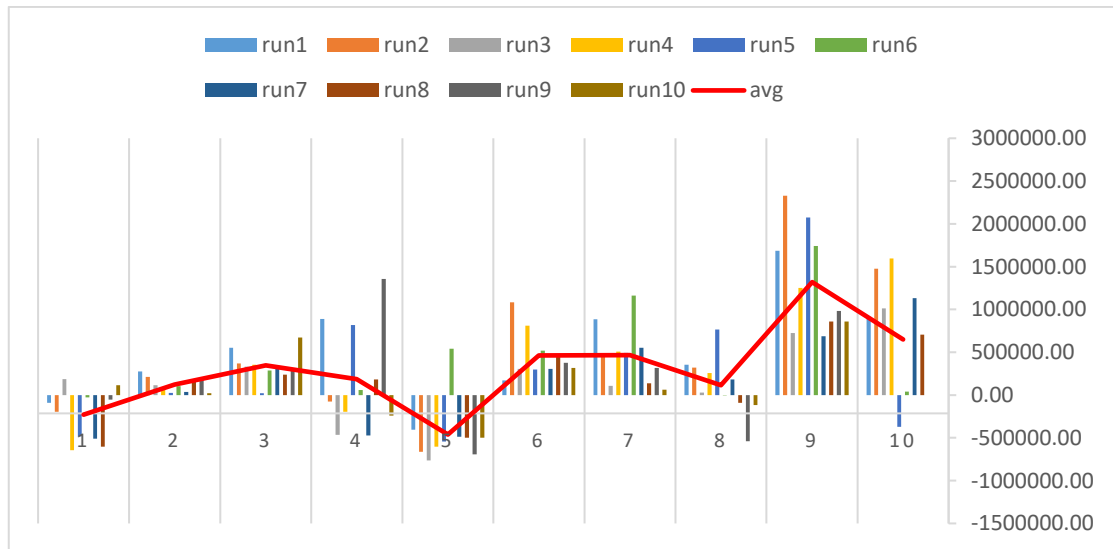
本節以選股模型（基本面）、純 GEP 交易模型（技術面）與混合應用模型（基本面+技術面）三者間的累計損益進行比較，結果如圖二十五所示。圖中顯示最佳表現為使用動態投組的 GEP 交易模型；最差則為只使用技術面的 GEP 模型。此外，不管是在使用 RF 或 LGBM 投組後，其獲利表現是遠勝於單純只使用 GEP 交易模型的方法。而從選股模型的買入持有策略進行比較，能發現單使用選股的方式就足以產生不錯的獲利，但加上 GEP 交易模型後有更進一步的提升，其中又以 RF 產生的投組表現最佳。

從各期表現來看能注意到，在經過第 5 與第 8 期後，除使用動態 RF 投組的策略外，其餘策略表現皆有明顯的拉回，且不只是在 GEP 交易模型能發現此一現象，在兩種決策樹選股模型中亦是如此。而從各期精確率結果也能注意到，該兩期的表現相較其他期有較明顯的落差。鑒於選股模型的精確率多少會影響到 GEP 交易模型的表現，然而，在一些精確率較高之情況下，GEP 交易模型的表現也並未皆有提升，其中原因可能是由於投組數量的限制影響。由於本研究只選取上漲標的中的前五名進行交易，故有一定機會選擇到剛好實際未上漲的股票。



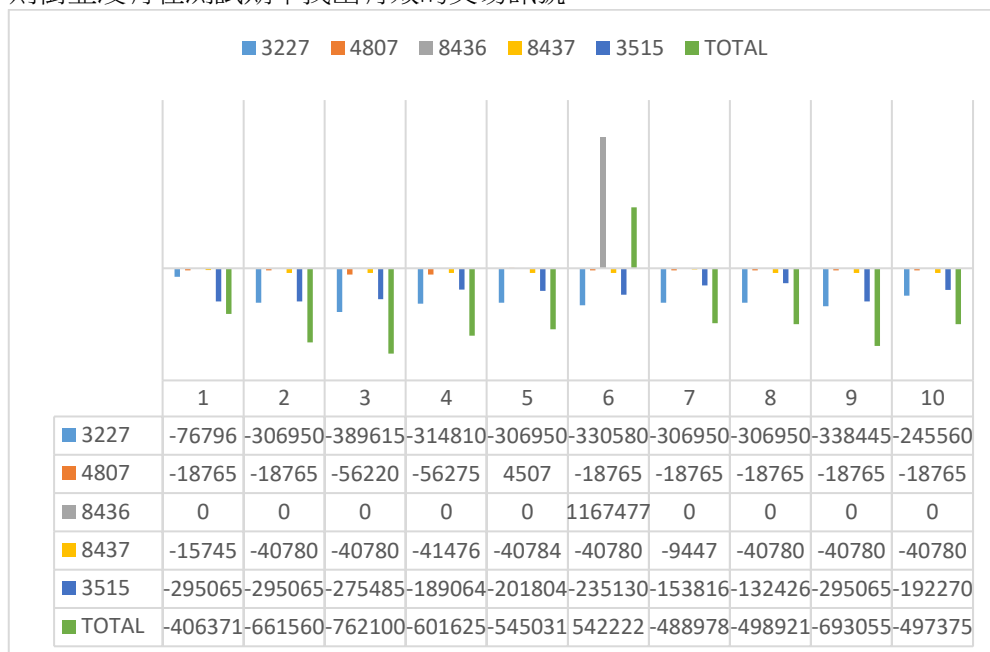
圖二十五 實驗三累計損益比較圖

總結上述綜合分析績效比較結果中能明確發現，GEP 交易模型於各期間的交易表現變化，在最佳的動態 RF 投組策略與動態 RF 投組策略平均相比下，GEP 模型的交易績效並不穩定。圖二十六顯示了 LGBM 動態投組十次實驗結果的交易損益情形，於各期損益的變化能看出在某些時期，GEP 的表現並不一致。以第五期為例，僅有一次的實驗在當期能獲利，而其餘九次的結果皆為虧損。



圖二十六 LGBM 動態投組各期損益表現示意圖

進一步觀察圖二十七，十次實驗中第五期的交易損益來看，除了 8436 這檔股票外，其餘各股表現相差並不大。且除了第六次實驗外，其他實驗中該檔股票的損益皆為零，代表在第五期中，8436 這檔股票的多數實驗所產生的規則樹並沒有在測試期中找出有效的交易訊號。



圖二十七 LGBM 動態投組十次實驗第五期損益示意圖

伍·結論與建議

一、結論

本研究以決策樹搭配基因表達規劃法建立一套動態調整投資組合的交易模型，在選股方面以 RF 與 LGBM 兩種決策樹透過六種選股因子作為特徵，預測 T+2 期的上漲標的，將每一期所預測之前五名上漲股票作為投資組合。在擇時部分則藉由基因表達規劃法中的交易訊號基因，以 7 種技術指標找出最合適之買賣點，並在每次買賣訊號出現時找出當下最適合之買賣張數，也透過風險管理基因找出策略之最佳停損停利點。根據本研究的實驗結果能得到以下結論：

1. 單以使用技術指標作為交易模型的 GEP 方法並無法獲得正報酬，且從每期表現來看，能發現其獲利情況浮動很大，表示在沒有選股的情況下，以多基因編碼結合交易訊號、資金管理與風險控管三種基因設計的方式，並非一穩定的交易策略。
2. 從實驗一與二的比較結果能發現，結合了動態選股的 GEP 交易模型確實能夠提升模型整體的績效，以總績效來看，不管在獲利、風險與總交易方面其表現皆是勝過純 GEP 交易模型。而從各期的績效表現來看，實驗一中兩種投組的第 5 與 8 期皆產生巨大虧損，而在實驗二中兩種投組在同一個時期卻有不同的表現，雖然在第 5 期時兩者皆為虧損但與實驗一的同期相比明顯降低不少，而在第 8 期 RF 投組的交易模型中，其交易報酬是由負轉正且贏過實驗一的同期模型。證實透過其他模型方法輔助 GEP 是可行的。
3. 在只使用選股模型的買入持有策略，就足以產生不錯的報酬結果。以混合模型的平均表現來看，不管是使用 RF 或是 LGBM 投組的交易策略，其表現皆與單獨使用選股模型的交易模式相差不大，但若是單以混合模型最佳的實驗結果來看，確實使用動態選股的方式能提高選股模型的報酬。
4. GEP 多基因編碼的表達樹，其所產生的交易規則至少有 3 種以上，但在測試期能發現交易次數的數量偏少，代表表達樹中僅有一至兩種規則被實際交易於測試期，其餘規則可能為無效規則。換句話說 GEP 在

訓練期中只找到了一條能夠獲利的策略，相反表達樹中其餘的策略並無經過適應函數評估，導致無效規則的產生。

二、研究貢獻

過去使用移動窗格訓練 GEP 動態交易之研究中，多數只專注於如何提升交易模型自身表現，往往忽略趨勢走向重要性，進而使交易產生龐大虧損。本研究參考了 Hsu (2021) 在選股模型所使用之六個因子，並以 GEP 作為擇時模型結合決策樹演算法來提升整體績效表現，分析未來具上漲潛力的股票作為投組。綜合上述，本研究有以下學術貢獻：

1. 在選股方面，本研究結合了黃賜傳 (2018) 於 GEP 交易模型中提出之動態標竿設計，陳姿穎 (2016) 之技術指標調整技術，以及李家瑋 (2020) 以 GEP 作為特徵工程篩選技術指標的研究，除了加入 GEP 交易方法於原有的選股策略，進一步，透過技術指標與動態天期找出個股每一時期最佳的合適買賣點，相較於過往以基本面分析找出長期趨勢的走向，能使策略更加的靈活並放大原有的績效表現。
2. 在 GEP 交易模型方面，本研究改良呂冠融 (2018) 所設計之 GEP 交易基因，將頭部長度由 4 擴展至 6，以及將屬性節點由四種技術指標增加至六種，提升 GEP 交易模型最大限制規則長度外並擴展交易規則搜尋範圍，並使用兩種決策樹模型結合選股因子預測未來兩季 ($T+2$) 具有上漲可能的股票，解決過去 GEP 動態交易方式原有的弱點。
3. 在交易方面，本研究同時考量了基本面與技術面，並應用李佳儒 (2010) 所提出之移動窗格方法模擬實際交易的情況，再透過機器學習模型進行選股與擇時的操作。過往許多研究僅考慮基本面 (Nti et al., 2019) 或單使用技術面 (Padhi et al., 2021)，且對市值龐大且穩定的股票進行的 GEP 實驗研究。而本研究採用決策樹選股方法，從台灣上市上櫃的公司中挑選標的進行實驗，相比過去靜態的單商品與單策略分析方法，多商品多策略的 GEP 交易模型，其實驗結果更能解釋 GEP 交易模型在不同商品不同時期的績效表現。

三、研究限制

本研究雖解決了動態 GEP 交易模型原有的一些弱點，但在研究結果上仍存在一些限制，分述如下：

1. 本研究之 GEP 交易模型僅使用過去相關研究中，常用的六種技術指標作為屬性節點。雖然以 GEP 的特性能透過 K 表達找出有效基因，然而，過少的指標可能仍會影響 GEP 搜尋的能力。
2. 在交易股票時包含做多 (long) 與放空 (short) 兩種操作方法，本研究之 GEP 交易模型單只設計做多（即低買高賣）的交易方式。但股市可能有牛市、盤整與熊市三種不同的盤勢，單以做多作為唯一交易方式可能限制了 GEP 交易模型的獲利能力。
3. 本研究結果發現使用多基因編碼方式，如交易次數過少，GEP 模型容易在訓練期找到一條獲利策略後就忽略了表達樹中其他的交易規則。由於本研究所設計適應函數僅評估觸發訊號之規則，有可能因訓練期的長度又或是交易標的價格波動，造成此問題產生。

四、未來研究

選股模型的精確率對於 GEP 交易模型之影響極大，鑑於此，未來研究在選股方面可針對三個方向改進。其一，找出不同的貼標方式使得模型在判別盤勢時更為準確，像是找出期間內最高價至最低價的落差，判別在此期間內是否出現過行情；其二，調整不同的模型參數，以最佳參數提升選股模型預測效果；其三，增加更多樣的特徵，如長線的技術指標或是籌碼指標。而在 GEP 方面，交易方式亦可增加做空的操作以擴展 GEP 在搜尋獲利策略的可能性，而績效指標能使用包含風險的指標作為目標，如夏普指標和風險報酬比等，期望使整體策略所獲得的權益更加平穩。

參考文獻

- 呂冠融，「深度學習與基因表達規劃法在台股智能交易系統之設計」，輔仁大學資訊管理學系碩士論文，2018 年。
- 李佳儒（譯），「交易策略評估與最佳化（第二版）」。台北市：寰宇出版社，2010 年。
- 李家瑋，「基於 C-RNN-GAN 神經網路的股票價格趨勢預測模型之研究-以美國股票市場為例」，輔仁大學資訊管理學系碩士論文，2020 年。
- 韋承佑，「基因表達規劃法為基的決策樹模型於股票動態交易策略發展之研究」，輔仁大學資訊管理學系碩士論文，2014 年。
- 張清良，「股票市場買賣研判指標的應用」，國立中正大學財務金融研究所碩士論文，2008 年。
- 張朝棟，「台股價值選股在 MA 與 MACD 操作策略之績效研究」，中華大學企業管理學系碩士論文，2011 年。
- 陳姿穎，「模糊基因表達規劃法在台股指數期貨之資金管理策略研究」，輔仁大學資訊管理學系碩士論文，2016 年。
- 陳帝豪，「基因表達規劃法為基的集成擇時交易策略之探勘」，輔仁大學資訊管理學系碩士論文，2012 年。
- 黃賜傳，「基因表達規劃演算法結合動態標竿在股票市場之研究」，輔仁大學資訊管理學系碩士論文，2018 年。
- 詹雅婷，「GEP-GA 為基的智慧型投資組合系統之設計」，輔仁大學資訊管理學系碩士論文，2015 年。
- 鍾祥龍，「毛利成長率、營收成長率與台股動量策略」，輔仁大學金融研究所碩士論文，2008 年。
- Allen, F., & Karjalainen, R. (1999). "Using genetic algorithms to find technical trading rules." *Journal of Financial Economics*, 51(2), 245-271.
- Barber, B., & Odean, T. (2000). "Trading is hazardous to your wealth: The common stock investment performance of individual investors." *The Journal of Finance*, 55, 773-806.
- Breiman, L. (2001). Random Forests. *Machine learning*, 45, 5-32.
- Bustos, O., & Pomares-Quimbaya, A. (2020). "Stock market movement forecast: A Systematic review." *Expert Systems with Applications*, 156, 113464.
- Chen, Y., Liu, K., Xie, Y., & Hu, M. (2020). "Financial trading strategy system based on machine learning." *Mathematical Problems in Engineering*, 2020(1), 1-13.
- Fama, E., & French, K. (2015). "A five-factor asset pricing model." *Journal of Finance Economic*, 116(1), 1-22.
- Fama, E. (1970). "Efficient capital markets: A review of theory and empirical work." *The Journal of Finance*, 25(2), 383-417.

- Ferreira, C. (2001). "Gene expression programming: A new adaptive algorithm for solving problems." *Complex Systems*, 13(2), 87-129.
- Ferreira, C. (2006). "Automatically defined functions in gene expression programming." In *Genetic Systems Programming: Theory and Experiences* (pp. 21-56). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hsu, T. Y. (2021). "Machine learning applied to stock index performance enhancement." *Journal of Banking and Financial Technology*, 5, 21-33.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). "LightGBM: A highly efficient gradient boosting decision tree." *Advances in Neural Information Processing Systems*, 30, 3146-3154.
- Khaidem, L., Saha, S., & Dey, S. R. (2016). "Predicting the direction of stock market prices using random forest." *Applied Mathematical Finance*, 1(5), 1-20.
- Ko, K. C., Lin, S. J., Su H. J., & Chang, H. H. (2014). "Value investing and technical analysis in Taiwan stock market." *Pacific-Basin Finance Journal*, 26, 14-36.
- Lekhal, M., & El Oubani, A. (2020). "Does the adaptive market hypothesis explain the evolution of emerging markets efficiency? Evidence from the Moroccan financial market." *Heliyon*, 6(7), e04429.
- Lim, K. P., & Brooks, R. (2011). "The evolution of stock market efficiency over time: A survey of empirical literature." *Journal of Economic Surveys*, 25(1), 69-108.
- Lo, A. W. (2004). "The adaptive markets hypothesis: Market efficiency from an evolutionary perspective." *Journal of Portfolio Management*, 30(5), 15-29.
- Michell, K., & Kristjanpoller, W. (2020). "Strongly-typed genetic programming and fuzzy inference system: An embedded approach to model and generate trading rules." *Applied Soft Computing*, 90, 106169.
- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2019). "A systematic review of fundamental and technical analysis of stock market predictions." *Artificial Intelligence Review*, 53(4), 3007-3057.
- Padhi, D. K., Padhy, N., Bhoi, A. K., Shafi, J., & Ijaz, M. F. (2021). "A fusion framework for forecasting financial market direction using enhanced ensemble models and technical indicators." *Mathematics*, 9(21), 2646.
- Rundo, F., Trenta, F., Di Stallo, A. L., & Battiato, S. (2019). "Machine learning for quantitative finance applications: A survey." *Applied Sciences*, 9(24), 5574.
- Sun, X., Liu, M., & Sima, Z. (2018). "A novel cryptocurrency price trend forecasting model based on LightGBM." *Finance Research Letters*, 32, 101084.
- Tan, Z., Yan, Z., & Zhu, G. (2019). "Stock selection with random forest: An exploitation of excess return in the Chinese stock market." *Heliyon*, 5(8), e02310.
- Urquhart, A., & Hudson, R. (2013). "Efficient or adaptive markets? Evidence from major stock markets using very long run historic data." *International Review of Financial Analysis*, 28, 130-142.
- Zhong, J., Feng, L., & Ong, Y. S. (2017). "Gene expression programming: A survey." *IEEE Computational Intelligence Magazine*, 12(3), 54-72.

- Zhong, J., Ong, Y. S., & Cai. W. (2016). "Self-Learning gene expression programming." *IEEE Transactions on Evolutionary Computation*, 20(1), 65-80.

A Study of Dynamic Trading Strategy Design Based on Gene Expression Programming-Evidence from the Taiwan Stock Market

QING-FENG LIN, SHIANG-LIN LIN, MING-YU LEE *

ABSTRACT

With the development of financial technology and artificial intelligence being applied to various financial products, it is a challenge to find a set of profitable trading strategies from the complex and ever-changing financial market. This study uses Gene Expression Programming (GEP) to search for effective trading strategies from past historical data, and builds a set of trading models based on three genes, namely, trading signals, capital management and risk control. In addition, two decision tree methods are used to dynamically update the investment portfolio in each period, namely, Random Forest and LGBM. Hoping through these two methods, the profit of the GEP trading model could be improved, and the risk of trading could be reduced. The results of this research found that combining dynamic stock selection with the GEP trading model, it can indeed effectively improve the trading performance of each period of the GEP trading model. Compared with the pure GEP model (technical aspect) and stock selection model (fundamental aspect), it can indeed effectively improve the trading performance of the GEP trading model in each period, such as profit, winning rate and risk.

Keywords : Random Forest, LGBM, Gene Expression Programming, Trade Strategy

* Qing-Feng Lin, Assistant Professor, Department of Information Management, Fu Jen Catholic University. Shiang-Lin Lin, Assistant Professor, Department of Information Management, Fu Jen Catholic University. Corresponding Author. Ming-Yu Lee, Master, Department of Information Management, Fu Jen Catholic University.